§103 rejections

In the 17 June 2008 Office Action claims 175 - 197 are rejected under §103 as being obvious given U.S. Patent 6,012,053 (hereinafter, Pant) in view of U.S. Patent 5,812,988 (hereinafter, Sandretto). The Examiner has cited Pant and Sandretto as references. The Assignee respectfully traverses the rejections for obviousness in several ways. First, by noting that the claim rejections are not in compliance with the Administrative Procedures Act and are therefore moot. Second, by noting the claim rejections are non-statutory. Third, by noting that the Office Action has failed to establish a prima facie case of obviousness.

In particular the cited combination of documents fails to establish a prima facie case of obviousness for claims 125 – 150 because the cited combination does not teach or suggest one or more of the limitations for every rejected claim. *MPEP 2143.03 provides that: to establish prima facie obviousness of a claimed invention, all the claim limitations must be taught or suggested by the prior art (In re Royka, 490 F.2d 981, 180 USPQ 580 (CCPA 1974)).* One limitation missing from the cited combination is impact which affects all claims.

§ 101 rejections

In the 17 June 2008 office action, claims 175 - 197 are rejected under 35 U.S.C. §101 as being unpatentable because the Examiner alleges that the disclosed invention lacks patentable utility as keyword relevancy to a corporation can allegedly not be determined. The Assignee will respectfully traverse the rejection of claims 175 - 197 under §101 in two ways. First, by noting that the Examiner has failed to establish a prima facie case of non utility. Second, by noting that the assertions regarding the alleged lack of utility are not in compliance with the requirements of the Administrative Procedures Act and are therefore moot. Third by noting that the claimed inventions transform transaction and text data into a different state or thing.

As mentioned previously, the Examiner has failed to establish a prima facie case of non utility for rejected claims 175 - 197. *MPEP 2164.07 states "the examiner has the initial burden of challenging an asserted utility. Only after the examiner has provided evidence showing that one of ordinary skill in the art would reasonably doubt the asserted utility does the burden shift to the applicant to provide rebuttal evidence sufficient to convince one of ordinary skill in the art of the invention's asserted utility. In re Brana, 51 F.3d 1560, 1566, 34 USPQ2d 1436, 1441 (Fed. Cir. 1995) (citing In re Bundy, 642 F.2d 430, 433, 209 USPQ 48, 51 (CCPA 1981)). The*

Assignee respectfully submits that the Examiner has not provided any evidence to support his assertions. Instead, he has simply made a series of arbitrary and capricious statements.

As noted previously, the second way the Assignee will traverse the § 101 rejections of claims 175 - 197 is by noting that the assertions regarding the alleged lack of utility are not in compliance with the requirements of the Administrative Procedures Act and are therefore moot. In Dickinson v. Zurko, 119 S. Ct. 1816, 50 USPQ2d 1930 (1999), the Supreme Court held that the appropriate standard of review of USPTO findings of fact are the standards set forth in the Administrative Procedure Act ("APA") at 5 U.S.C. 706 (1994). The APA provides two standards for review – an arbitrary and capricious standard and a substantial evidence standard. The Assignee respectfully submits that the 35 U.S.C. § 101 rejection of claims 175 - 197 in the instant Office Action fails under both standards. It fails under the substantial evidence standard because as detailed above no evidence was presented. The Examiner is undoubtedly unwilling to make any specific allegations because all of the specific allegations made in the past have proved to be incorrect. It also fails under the arbitrary and capricious standard because the U.S.P.T.O. has found claims for determining keyword relevance to be patentable in the case of the Pant patent and for a copending application a few months ago.

§ 112 first paragraph rejections

In the 17 June 2008 Office Action the Examiner has rejected claims 175 - 197 under 35 U.S.C. §112 first paragraph as lacking a written description that would enable those of average skill in the art to make and use the claimed invention. Specifically, the Examiner has made an unsupported allegation that the invention requires subjective judgments and is by definition unique, arbitrary and subjective.

The Assignee will respectfully traverses the §112 first paragraph rejection of claims 175 - 197 in three ways. First, by noting that the Office Action has failed to establish a prima facie case that the specification does not meet the requirements of §112 first paragraph. Second, by noting that the assertions regarding the alleged lack of written description are not in compliance with the both standards of the Administrative Procedures Act and are therefore moot. Third by noting the rejections are non-statutory.

As mentioned previously, the Examiner has failed to establish a prima facie case that the specification does not meet the requirements of §112 first paragraph. MPEP 2163 states that:

"A description as filed is presumed to be adequate, unless or until sufficient evidence or reasoning to the contrary has been presented by the examiner to rebut the presumption. See, e.g., In re Marzocchi, 439 F.2d 220, 224, 169 USPQ 367, 370 (CCPA 1971). The examiner, therefore, must have a reasonable basis to challenge the adequacy of the written description. The examiner has the initial burden of presenting by a preponderance of evidence why a person skilled in the art would not recognize in an applicant's disclosure a description of the invention defined by the claims. Wertheim, 541 F.2d at 263, 191 USPQ at 97. In rejecting a claim, the examiner must set forth express findings of fact regarding the above analysis which support the lack of written description conclusion. These findings should:
    (A) Identify the claim limitation at issue; and
    (B) Establish a prima facie case by providing reasons why a person skilled in the art at the time the application was filed would not have recognized that the inventor was in possession of the invention as claimed in view of the disclosure of the application as filed. A general allegation of "unpredictability in the art" is not a sufficient reason to support a rejection for lack of adequate written description."

The arguments presented by the Examiner fail to establish the prima facie case required to sustain a §112 first paragraph rejection for a single claim in at least three ways:

1. the first way the 17 June 2008 Office Action fails to establish the prima facie case that the specification does not meet the requirements of §112 first paragraph is that the Examiner has not identified any reasons why a person skilled in the art at the time the application was filed would not have recognized that the inventor was in possession of the invention as claimed. To the contrary, there is substantial evidence that the comments regarding the written description were provided by individuals and an organization that lacks the necessary skill in the relevant arts;

2. the second way the 17 June 2008 Office Action fails to establish the prima facie case that the specification does not meet the requirements of §112 first paragraph is that the Examiner has only made a general allegations (See prior 35 U.S.C. § 101 Rejection of Claims discussion); and

3. the third way the 17 June 2008 Office Action fails to establish a prima facie case that the specification does not meet the requirements of §112 first paragraph is that the Examiner has not identified the claim limitation(s) at issue.

The Assignee respectfully submits that the assertion that the specification does not meet the requirements of §112 first paragraph also fail under both standards of the APA. First, as detailed above, the Examiner has not provided any evidence to support these allegations. As a result, the §112 first paragraph rejection of claims 175 - 197 fails under the substantial evidence standard. The Examiner is undoubtedly unwilling to make any specific allegations because all of the specific allegations made in the past have proved to be incorrect. Second, a comparison of the method disclosed in the instant application with the description contained in a recently

issued patent shows that the rejections fail under the arbitrary and capricious standard. This is because instant application has no identifiable subjectivity in model development while considerable subjectivity appears to be present in a recently issued patent for completing similar tasks:

Summary of 10/750,792	Summary of 7,283,982 filed in 2003
1. Transform raw data into indicators using pre-programmed functions and Linus/AQ algorithms	1. Use <u>any</u> technique to derive a basic model

2. Develop an initial model using the raw and transformed data as inputs by:
   a) creating parallel models using different specified algorithms,
   b) using **stepwise regression** to identify the best set of input variables for the models for each algorithm type, and
   c) selecting the three best algorithm type models from b)
	2. Develop an initial model by:
	   a) deriving features from the input to the basic model using <u>any</u> current transform regression algorithm, and
	   b) using **stepwise regression** to select the input features for the initial regression model

3. Refine the variable selection from 2b) for the three best models from 2c) and then transform the resulting set of input variables into summaries using different specified algorithms.	3. Complete a non-linear transformation of an explanatory input feature(s) from the initial model.
	4. Use the transformed input features to create a new linear regression model
4. Use the best summary of transformed data from 3 to create a final regression model	5. Combine the output of the new linear regression model with the output of the initial model and use the sum to provide a final model for the current iteration
	6. Repeat steps 3 through 5 indefinitely

As shown above, both the instant application and issued patent methods rely on stepwise regression for the input variable selection step. The use of stepwise regression for variable selection has been judged to be obvious (see 103 rejection discussion) and/or too subjective in the instant application and in related applications that develop models (it was equated with the use of fear and emotion in an Office Action for application 09/688,983) <u>while the issued patents reliance on the exact same technique was judged to be novel and concrete when used in model development</u>.

Furthermore, the written description for the issued patent appears to be more subjective than the written description for the Asset Reliance application because: one step in the issued patent

method calls for the use of "any" model, another step in the issued patent method calls for the use of "any" current transform regression algorithm, and the number of iterations is not specified. By way of contrast, the instant application specifically identifies the algorithms used at every step of the model development process and does not use an open ended model development process.

While no rebuttal is required, the Assignee also notes that a declaration has been provided which also could be used to provide a complete rebuttal of the unsubstantiated allegations contained in the 17 June 2008 Office Action regarding a lack of written description.

§ 112 Second Paragraph Rejections

In the 17 June 2008 Office Action the Examiner has rejected claims 175 - 197 under 35 U.S.C. §112 first paragraph as lacking a written description that would enable those of average skill in the art to make and use the claimed invention.

The Assignee will respectfully traverses the §112 second paragraph rejection of claims 175 - 197 in three ways. First, by noting that the Office Action has failed to establish a prima facie case that the specification does not meet the requirements of §112 second paragraph. Second, by noting that the assertions regarding the alleged lack of written description are not in compliance with the both standards of the Administrative Procedures Act and are therefore moot. Third by noting the rejections are non-statutory.

As mentioned previously, the first way the Assignee will traverse the 35 U.S.C. §112 second paragraph rejection of claims 175 - 197 will be by noting that the arguments presented by the Examiner fail to establish the prima facie case required to sustain a §112 second paragraph rejection. *MPEP 2173.02 states that: definiteness of claim language must be analyzed, not in a vacuum, but in light of:*

(A) The content of the particular application disclosure;
(B) The teachings of the prior art; and
(C) The claim interpretation that would be given by one possessing the ordinary level of skill in the pertinent art at the time the invention was made.

In reviewing a claim for compliance with 35 U.S.C. 112, second paragraph, the examiner must consider the claim as a whole to determine whether the claim apprises one of ordinary skill in the art of its scope and, therefore, serves the notice function required by 35 U.S.C. 112, second paragraph, by providing clear warning to others as to what constitutes infringement of the patent. See, e.g., Solomon v. Kimberly-Clark Corp., 216 F.3d 1372, 1379, 55 USPQ2d

1279, 1283 (Fed. Cir. 2000). See also In re Larsen, No. 01-1092 (Fed. Cir. May 9, 2001). In the case of claims 175 - 197 the Examiner has failed to establish the prima facie case that the specification does not meet the requirements of §112 second paragraph in four ways for every rejected claim. The four ways are:

1. by failing to interpret the claims in light of the specification,

2. by failing to provide any evidence that someone of average skill in the relevant arts would have difficulty interpreting the claims,

3. by failing to establish that the limitation(s) in the claims fail to describe the invention and/or

4. by failing to consider the claim as a whole.

The detail cited under the discussion of the §112 first paragraph rejection discussion of failure to comply with the APA also supports the arguments regarding the APA under this section.

Statement under 37 CFR 1.111

37 CFR 1.111 requires that the basis for amendments to the claims be pointed out after consideration of the references cited or the objections made. 37 CFR 1.111 states in part that:

> In amending in response to a rejection of claims in an application or patent undergoing reexamination, the applicant or patent owner must clearly point out the patentable novelty which he or she thinks the claims present in view of the state of the art disclosed by the references cited or the objections made. He or she must also show how the amendments avoid such references or objections.

The Assignee notes that this requirement is not relevant to the instant application because, as detailed above, there are no references or objections to avoid. Having said that, the Assignee notes that the primary reasons the prior set of claims were amended to put the claims in final form for allowance and issue.

Request for affidavits under 37 C.F.R. 1.104

Because the 17 June 2008 Office Action contains no evidence, the claim rejections rely entirely on the personal knowledge of the Examiner and/or one or more other employees of the Office. 37 C.F.R. 1.104 provides that:

> *When a rejection in an application is based on facts within the personal knowledge of an employee of the Office, the data shall be as specific as possible, and the reference must be supported, when called for by the applicant, by the affidavit of such employee, and*

such affidavit shall be subject to contradiction or explanation by the affidavits of the applicant and other persons.

Accordingly, the Assignee requests that affidavits detailing the facts within the personal knowledge of any employee(s) of the Office that <u>address</u> explains:

1. ~~why using the tangible impact on cash flow to determine keyword relevancy is arbitrary and indefinite – the affidavit should address why Accenture (see Ballow reference), Asset Economics (see Ballow reference), Ernst & Young (see Measures that Matter) and others feel that identifying such impacts is possible – in short where are they wrong?,~~

2. ~~why using the tangible impact on corporate financial performance on corporate financial performance to determine keyword relevancy is arbitrary and indefinite – the affidavit should address why Accenture (see Ballow reference), Asset Economics (see Ballow reference), Ernst & Young (see Measures that Matter) and others feel that identifying such impacts is possible – in short where are they wrong?,~~

3. ~~why no solution to the problem of keyword relevancy is possible, the affidavit should address how the Pant invention, Google's search (www.google.com) and page rank algorithm (U.S. Patent 6,285,999) and Yahoo (search.yahoo.com) appear to have overcome this unsolvable problem;~~

4. ~~why similar methods for developing models to determine relationships are considered concrete and statutory when developed by a large company and subjective and obvious when disclosed by the Assignee 3 years earlier;~~

5. ~~why the Examiner feels that subjective inputs are required, the identify of the subjective inputs required for the operation of the claimed invention that allegedly affect the final results and what affect a mis-specification of any the subjective inputs that allegedly affect the final result have on the results produced by the claimed invention;~~

6. ~~why the use of subjective inputs is non-statutory (see for example Tulskie 6,249,768) and/or the fact that results are going to be reviewed subjectively renders an invention non-statutory (cite relevant statute or precedent),~~

7. ~~why an almost identical method for data integration is considered concrete and novel when introduced by a large company and subjective and obvious when disclosed by the Assignee almost 2 years earlier,~~

8. ~~why large companies are allowed to describe inventions for integrating data in accordance with up to 3 schemas without explaining any of them (see Appendix),~~

9. ~~why the Examiner previously argued that integrating data in accordance with a common schema was old and well known in the art and was implicit in a number of references,~~

10. ~~why the use of stepwise regression to select the variables used in models is considered subjective and obvious when disclosed by the Assignee for use in financial performance management while being considered concrete and novel when disclosed by a large company 4 years later for~~ any domain.

1. The clearly erroneous assertion that using the tangible impact on cash flow to determine keyword relevance is arbitrary and indefinite. For example Accenture (see Ballow reference), Asset Economics (see Ballow reference) and Ernst & Young (see Measures that Matter) teach that concretely measuring similar impacts is possible. Please provide an affidavit detailing why using the tangible impact on cash flow to determine keyword relevancy is arbitrary and indefinite – the affidavit should also address why the others that feel that concretely identifying such impacts is possible are incorrect,

2. The clearly erroneous assertion that using the tangible impact on corporate financial performance to determine keyword relevance is arbitrary and indefinite. For example Accenture (see Ballow reference), Asset Economics (see Ballow reference) and Ernst & Young (see Measures that Matter) teach that concretely measuring similar impacts is possible. Please provide an affidavit detailing why using the tangible impact on corporate financial performance to determine keyword relevancy is arbitrary and indefinite – the affidavit should also address why the others that feel that concretely identifying such impacts is possible are incorrect,

3. The assertion that no solution to the problem of keyword relevancy is possible which is clearly in error in light of the Pant invention and the fact that Google's search (www.google.com) and Yahoo (search.yahoo.com) appear to have made useful strides in overcoming this problem. Please provide an affidavit detailing the reasons why a solution to the keyword relevancy problem is not possible in light of issued patents and the apparent success of others in addressing this problem,

4. The erroneous assertion that establishing the common schema used in the instant application is not possible or well explained. This assertion is traversed by noting that establishing said schema requires only that a user identify the account numbers

associated with the components of value and real options, identify unit of measure conversions and identify the data associated with each element of value (note: many elements of value have their own systems) using a metadata mapping and conversion rules window and by noting that Warshavsky (6,732,095) uses a very similar method for identifying the dtd that is used for data conversion and it is considered concrete, Lyons (4,989,141) teaches that users have the requisite knowledge to complete this task as it allows all of them to use this knowledge create their own schemas and that Tulskie (6,249,768) and Harhen (5,406,477) also teach that users have the requisite knowledge as they teach that users can use this knowledge to invent data for use in planning and valuation analyses. The assertion regarding the alleged deficiency is also traversed by noting that the use of schemas is apparently so well known in the art that large companies are allowed to describe inventions for integrating data in accordance with up to 3 schemas without explaining any of them (see Appendix). Furthermore, the Examiner and others in TC 3600 have previously argued that integrating data in accordance with a common schema was old and well known in the art (see cited Office Actions). Please detail the alleged difficulties associated with establishing a common schema in light of the prior art, prior statements and the cited patents, and please describe the level of skill of the person of average skill in the art that would allegedly have this difficulty as part of said affidavit.

5. A review of the table shown under the 112 first paragraph discussion clearly shows that the Examiner's assertion that subjective inputs affect the final results and that a mis-specification of any the subjective inputs will have an affect on the results produced by the claimed invention is in error. Please provide an affidavit detailing why the use of stepwise regression to select the input variables used in models is considered subjective and obvious when disclosed by the Assignee for use in financial performance management in the instant application while being considered concrete and novel when disclosed by a large company 4 years later for any domain. The affidavit should also address why IBM is incorrect in asserting that its use of the same method used in the instant application for input selection for any domain is not concrete or objective.

The Assignee also requests that an affidavit detailing the combinations and/or modifications of teachings that the Examiner and other personnel at the U.S.P.T.O. who will be involved in the review of this application, its anticipated continuation and the anticipated appeal have made without the assistance of a patent specification or any other teaching, motivation or suggestion

be prepared and forwarded to the Assignee. For each listed combination, the teaching(s) and their source should be identified. The product name for any commercialized combinations of teachings should also be included in the affidavit. If there are none, then the Assignee is also requesting an affidavit detailing the statutory basis for allowing individuals with a well documented lack of skill in the relevant arts to author 103 and/or 112 rejections. The Assignee hereby also requests an affidavit detailing the statutory basis for using different standards to review similar inventions based on company size.

Copending applications

The Assignee has a number of applications pending for search related inventions, however, none of them utilize the method outlined in the instant application. Application 10/645,099 has three claims for completing an internet search for keywords (claims 60 -62).

Under the provisions of MPEP § 2001.06(b), the Examiner is hereby advised of the following Office Actions for application that are and/or were copending that may be relevant include the 11/21/2000 Office Action for 08/999,245, the 10/20/2006 Office Action for 08/999,245, the 02/20/2007 Office Action for 08/999,245, the 09/10/2007 Office Action for 08/999,245, the 11/26/2001 Office Action for 09/293,336, the 01/06/2004 Office Action for 09/688,982, the 08/28/2006 Office Action for 09/688,982, the12/24/2003 Office Action for 09/688,983, the 06/15/2004 Office Action for 09/688,983, the 02/24/2005 Office Action for 09/688,983, the 03/09/2006 Office Action for 09/688,983, the 01/03/2007 Office Action for 09/688,983, the 10/11/2005 Office Action for 09/761,670, the 05/09/2006 Office Action for 09/761,670, the 01/03/2007 Office Action for 09/761,670, the 05/03/2004 Office Action for 09/761,671, the 11/18/2004 Office Action for 09/761,671, the 09/30/2005 Office Action for 09/761,671, the 06/13/2006 Office Action for 09/761,671, the 10/06/2004 Office Action for 09/764,068, the 07/01/2005 Office Action for 09/764,068, the 12/30/2005 Office Action for 09/764,068, the 07/12/2006 Office Action for 09/764,068, the 08/23/2007 Office Action for 09/764,068, the 05/23/2003 Office Action for 09/940,450, the 11/24/2003 Office Action for 09/940,450, the 05/03/2006 Office Action for 09/940,450, the 09/07/2006 Office Action for 10/012,375, the 02/27/2007 Office Action for 10/012,375, the 08/23/2007 Office Action for 10/012,375, the 02/25/2008 Office Action for 10/012,375, the 07/12/2006 Office Action for 10/025,794, the 11/14/2007 Office Action for 10/025,794, the 03/31/2008 Office Action for 10/025,794, the 02/14/2007 Office Action for 10/036,522, the 11/14/2007 Office Action for 10/036,522, the 02/27/2007 Office Action for 10/097,344, the 08/23/2007 Office Action for 10/097,344, the

06/26/2007 Office Action for 10/166,758, the 7/11/2007 Office Action for 10/237,021, the 2/06/2008 Office Action for 10/237,021, the 12/20/2004 Office Action for 10/282,113, the 10/20/2005 Office Action for 10/282,113, the 02/24/2006 Office Action for 10/282,113, the 05/26/2006 Office Action for 10/282,113, the 03/27/2008 Office Action for 10/287,586, the 08/08/2003 Office Action for 10/012,374, the 01/21/2004 Office Action for 10/012,374, the 05/24/2004 Office Action for 10/012,374, the 08/23/2004 Office Action for 10/012,374, the 07/06/2005 Office Action for 10/012,374, the 05/22/2006 Office Action for 10/012,374, the 04/09/2007 Office Action for 10/046,094, the 12/04/2007 Office Action for 10/046,094, the 08/11/2004 Office Action for 10/329,172, the 06/08/2005 Office Action for 10/329,172, the 12/02/2005 Office Action for 10/329,172, the 07/24/2006 Office Action for 10/329,172, the 12/16/2003 Office Action for 10/441,385, the 11/02/2005 Office Action for 10/441,385, the 06/02/2006 Office Action for 10/441,385, the 11/03/2006 Office Action for 10/441,385, the 02/27/2007 Office Action for 10/441,385, the 08/10/2007 Office Action for 10/441,385, the 07/02/2004 Office Action for 10/674,861, the 09/26/2006 Office Action for 10/645,099, the 02/27/2007 Office Action for 10/645,099. the 01/09/2008 Office Action for 10/645,099. the 02/06/2007 Office Action for 10/743,616, the 07/26/2007 Office Action for 10/743,616, the 02/11/2007 Office Action for 10/743,616, the 04/12/2005 Office Action for 10/746,673, the 07/10/2006 Office Action for 10/746,673, the 01/23/2007 Office Action for 10/746,673, the 02/27/2007 Office Action for 10/746,673, the 11/20/2007 Office Action for 10/746,673, the 02/14/2008 Office Action for 10/747,471, the 08/09/2007 Office Action for 10/821,504, the 6/27/2007 Office Action for 11/360,087 the 11/01/2007 Office Action for 11/360,087, the 6/17/2008 Office Action for 10/750,792, the 6/18/2008 Office Action for 11/278,419 and the 6/06/2008 Office Communication for 10/441,385. In almost all cases these Office Actions and Communications are available from the U.S.P.T.O. Internet File Wrapper system.

Reservation of rights

The Assignee hereby explicitly reserves the right to present the previously modified and/or canceled claims for re-examination in their original format. The cancellation or modification of pending claims to put the instant application in a final form for allowance and issue is not to be construed as a surrender of subject matters covered by the original claims before their cancellation or modification.

Conclusion

The pending claims are of a form and scope for allowance. Prompt notification thereof is respectfully requested.

Respectfully submitted,
Asset Trust, Inc.

/B.J. Bennett/

B.J. Bennett, President
Date: June 29, 2008

APPENDIX

to numerous problems in corporate and database mining such risk assessment, attrition and retention modeling, campaign marketing, fraud detection, customer profiling, profitability and cross-selling. These application problems are usually viewed from an account- or user-centric point of view. All the relevant information for each user is merged and consolidated in one record. An input dataset then looks like a large, mostly populated two-dimensional table where the columns correspond to attributes (independent variables). In the supervised learning approach, one particular column provides the 'target' that is used as the dependent variable for the Data Mining model. Association modeling attempts to find associations: common patterns and trends in a less structured way (i.e. independent of a particular target field). These associations are supported by statistical correlations between different attributes of the dataset and are extracted by imposing independence, support, and confidence thresholds. Association analysis is applied to transaction or market basket data typically. In this case the datasets consists of transaction data listing a basket or group of items corresponding to an individual sale. The dataset is again a two-dimensional table but in this case potentially very sparse. Clustering is used for data-reduction and for class discovery. It is a method to find general correlation structures that group records into similarity groups. Clustering can be applied to both account or transaction-based datasets. Most data mining tool-sets support algorithms that provide instances of these paradigms but it is not common to encounter the three paradigms in a single problem.

Enterprise web mining (EWM) in its most general realization involves a collection of data intensive data sources and repositories with corporate, warehousing and webtransaction components. As a consequence of this heterogeneity the present invention must incorporate these data sources in a way suitable to support the three learning paradigms and also allow the system to solve different types of mining problems along the spectrum of web enterprises shown in FIG. 3. On one side of the spectrum the present invention provides the capability to perform traditional data mining modeling on corporate RDBMS augmented by account-centric web data. For example, modeling of attrition in a phone company. On the other side of the spectrum the present invention provides the capability to perform pure transactional association analysis such as the one needed in sites such as search engines. Most web sites and corporate enterprises are somewhere in the middle.

Thus, the present invention provides the capability to

Extract session information from web server data.

Transform a web site visitor's behavior into data about his preferences.

Integrate web transactions and browsing behavior data with customer information and demographics

Support a variety of mining problems (e.g., cross-selling, up-selling, market segmentation, customer retention, and profitability) that use as input web and corporate data.

Help discover interesting and relevant patterns, clusters, and relationships in the transaction and user customer data.

An important function performed by the present invention is to integrate many existing information gathering, storage and decision elements in a coherent way. In order to do this, the methodology in the integration process and in the user interface must be defined.

It is useful to distinguish three types of web mining. Web mining consisting of web-deployed traditional data mining

provides the capability for web pages to use results of segmentation models for advertisements, cross-selling, etc. Web mining consisting of data-mining of click stream data provide the capability to generate statistical usage reports, on-line personalized recommendations, and on-line personalized navigation and general content. Full-fledged Enterprise Web Mining, as provided by the present invention, provides the capability to integrate traditional mining and click stream and conceptual classes encompassing the entire corporate/web customer life-cycle, including acquisition, cross-selling, and retention. In addition, it provides the capability to implement a dynamically personalized virtual store with artificial intelligence sales agents.

Another important aspect of the present invention is the personalization application. The personalization application is an integrated software application that provides a way for a Web site to customize—or personalize—the recommendations it presents to Web site visitors and customers.

Recommendations are personalized for each visitor to the Web site. This has distinct advantages over tailoring recommendations to broad, general market segments. Recommendations are based on a visitor's data and activity such as navigational behavior, ratings, purchases, as well as demographic data.

The personalization application collects the data and uses it to build predictive models that support personalized recommendations of the form "a person who has clicked links x and y and who has demographic characteristics a and b is likely to buy z".

The personalization application incorporates visitor activity into its recommendations in real time—during the Web visitor's session. For example, the personalization application records a visitor's navigation through the Web site, noting the links that are clicked, etc. All this is data stored for that visitor. The visitor may respond to a Web site's request to rate something, e.g., a book or a movie; the rating becomes part of the data stored for that visitor. All the Web-based behavior for the visitor is saved to a database, where the personalization application uses it to build predictive models. This data can be updated with data collected in subsequent sessions, thereby increasing the accuracy of predictions.

The personalization application works in conjunction with an existing Web application. The Web application asks the personalization application to record certain activities, and the data is saved by the personalization application into a schema. The Web application asks the personalization application to produce a list of products likely to be purchased by a Web site visitor; a scored list of recommendations compiled from the visitor's current behavior and from data in another schema is passed to the Web application. A third schema maintains administrative schedules and activities.

The personalization application collects four kinds of data:

navigational behavior

ratings

purchases

demographic data

Of these, navigational behavior allows the most flexibility. It can represent anything the Web application wants to consider a hit (e.g., viewing a page, clicking a link/item, etc.).

Visitors to the Web site are of two types: registered visitors (customers) and unregistered visitors (visitors). For customers, the personalization application has both data from a current session and historical data collected over time for a given customer, as well as demographic data. For